

The NIFTY way of Bayesian signal inference

Marco Selig

*Max Planck Institut für Astrophysik (Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany), and
Ludwig-Maximilians-Universität München (Geschwister-Scholl-Platz 1, D-80539 München, Germany)*

Abstract. We introduce NIFTY, “Numerical Information Field Theory”, a software package for the development of Bayesian signal inference algorithms that operate independently from any underlying spatial grid and its resolution. A large number of Bayesian and Maximum Entropy methods for 1D signal reconstruction, 2D imaging, as well as 3D tomography, appear formally similar, but one often finds individualized implementations that are neither flexible nor easily transferable. Signal inference in the framework of NIFTY can be done in an abstract way, such that algorithms, prototyped in 1D, can be applied to real world problems in higher-dimensional settings. NIFTY as a versatile library is applicable and already has been applied in 1D, 2D, 3D and spherical settings. A recent application is the D³PO algorithm targeting the non-trivial task of denoising, deconvolving, and decomposing photon observations in high energy astronomy.

INTRODUCTION

To learn from data is a fundamental concept in science and engineering. In many applications, we want to infer a physical continuous function – a “field” in physicists’ terms – from the measured data. In high energy astronomy, for example, X-ray observations are studied for the underlying photon flux distribution in order to deepen our understanding of the causal astrophysical processes. Likewise in the field of medical imaging, X-ray screenings revealing the physical matter density are prescribed due to their diagnostic power. Both, the photon flux and the matter density, are continuous functions of some position space, solid angle on the sky or (projected) spatial position. We denote such a function we are interested in as signal field.

Before we address the question of how to infer a signal field from a given data set, let us reflect on the conceptual differences between data and signal. Formally, data can be regarded as a finite set of numbers obtained from a measurement. For convenience, we condense all available data in a single data vector $d = (d_1, d_2, d_3, \dots)^T$. The dimension of this vector is finite since the measurement, for practical reasons, is of finite resolution and coverage with regard to position, time, energy, and the like. A signal field, however, might vary continuously with respect to those quantities. We assume the signal field $s = s(x)$ to be a continuous function on some space \mathcal{U} , with $x \in \mathcal{U}$. This space \mathcal{U} can describe a position, time, or energy domain; or even a combination of such domains. The signal field s is thereby defined for every $x \in \mathcal{U}$, and has infinitely many degrees of freedom. In (computational) practice, we need to discretize each field of course for usage within a computer algorithm.

The inverse problem of constructing an estimate for the signal field from a finite data set is, in general, ill-posed due to the profound signal ambiguity. That means there are (infinitely) many signal field configurations that can lead to the very same data. That is because, on the one hand, we lose information on the signal during the measurement process, where an infinite-dimensional quantity is mapped to a finite-dimensional one. For example, fluctuations in the photon flux can only be captured on spatial scales that the observation is able to resolve. On the other hand, the measurement is prone to noise; e.g., Poissonian shot noise in the case of integer photon counts. Even if a measurement of a (unchanged) signal is repeated under identical conditions, a different noise realization can lead to different data.

The literature on methodological approaches to solve such inverse problems is extensive. Most successful inference schemes are based on Bayesian reasoning using maximum entropy arguments and exploiting (empirically or theoretically motivated) assumptions on likelihood and prior statistics [1, 2, 3]. One theory, among many other, is Information Field Theory (IFT) [4, 5] that raises Bayesian signal inference to a statistical field theory. The goal hereby remains the derivation of optimal filters – in the sense of least square error – that applied to a data vector yield a posterior mean estimate for the signal field. A classical example would be the well-known Wiener filter [6], which in the IFT frame corresponds to a free (interactionless) theory.

Formally, we can derive and formulate inference algorithms in an abstract and mathematical language. Therefore, there is no need to explicitly specify the dimensionality of the problem, and in particular to specify neither the space

\mathcal{U} nor its counterpart in the computer environment. In theory, we could apply a worked out algorithm disregarding the resolution, the total size, and even the geometry of the inference scenario at hand. This kind of universality is part of the success story of inference algorithms like the Wiener filter, which is applied in the most diverse areas.

Why should we not transfer this universality of inference algorithms to their numerical implementations? The need of and benefit from multi-purpose algorithms that operate independently of the topology and dimensionality of the underlying spaces seems obvious. X-ray imaging techniques developed for high energy astronomy, for example, might also be applicable to medical imaging problems, and vice versa. This would also grant the advantage of prototyping inference algorithms in a simplified setting and of reusing existing codes. Notice that such an algorithm also yields signal field estimates that are independent of the chosen computational grid and its resolution. The reconstruction of a signal field with low and high resolution would look alike up to features that are only representable in the latter.

All this functionality has now become possible for users resorting to the NIFTY package. NIFTY abbreviates “Numerical Information Field Theory”, and is publicly available at <https://www.mpa-garching.mpg.de/ift/nifty/>.

NIFTY

The NIFTY library is written in PYTHON in order to provide the user with a clear and easily understandable interface. Another advantage of the PYTHON interface is that users can make use of the huge number of available packages. These comprise sophisticated numerical libraries as well as application oriented packages; e.g., for astronomical instruments. For computational heavy tasks, libraries written in CYTHON, C++ and C are used for efficiency.

The most important property of NIFTY is that it operates regardless of the underlying spatial grid and its resolution. Say we want to synthesize a statistically homogeneous and isotropic Gaussian random field from a (Fourier or angular) power spectrum, for example. Given a concrete power spectrum $p(k)$, we know which amplitudes of fluctuations to expect on certain (spatial, time or energy) scales $x \propto 1/k$. If we draw a random field from this power spectrum, we expect the same kind of fluctuations no matter how we define the space \mathcal{U} . In other words, we could choose \mathcal{U} to be a one- or two-dimensional grid with low or high resolution, but each random field realization would exhibit similar fluctuations.

NIFTY makes this possible through an object-orientated framework that offers, among others, classes for spaces, fields, and operators. Here, a space represents a certain type of computational grid. A field, which needs to be defined on some space, primarily carries an array of field values. Operators, which are applied to fields, perform specified (not necessarily linear) transformations. All grid dependent computations are hereby forwarded to the corresponding space, where the correct transformations and normalizations are executed.

In order to emphasize the implied subtlety, let us consider two scalar products. In the data domain, scalar products are defined as in any linear algebra textbook,

$$d \bullet a = \sum_i d_i^* \cdot a_i, \quad (1)$$

where $*$ denotes complex conjugation. The scalar product between signal fields, however, is described by an integral over the continuous space \mathcal{U} ,

$$s \bullet u = \int_{\mathcal{U}} dx \, s^*(x) \cdot u(x). \quad (2)$$

In order to approximate such integrals by (finite) sums in the computer environment, a proper discretization scheme for \mathcal{U} needs to be specified. For this purpose, we define a mapping, $s(x) \mapsto s_q$, that discretizes a signal field; e.g., by an evaluation at certain grid positions, $s_q = s(x_q)$. This yields

$$s \bullet u \approx \sum_q w_q \cdot s_q^* \cdot u_q, \quad (3)$$

where the summation respects the volume elements dx in form of an appropriately chosen weights w_q ; e.g., the physical volume of the considered bin, cell, pixel, or voxel indexed by q . Those weights are important for the proper normalization, not only of scalar products, but of all operations that involve integrations over \mathcal{U} . A direct consequence of this is that the continuum limit is always preserved. In this way, the resolution independence of the calculus with fields is ensured.

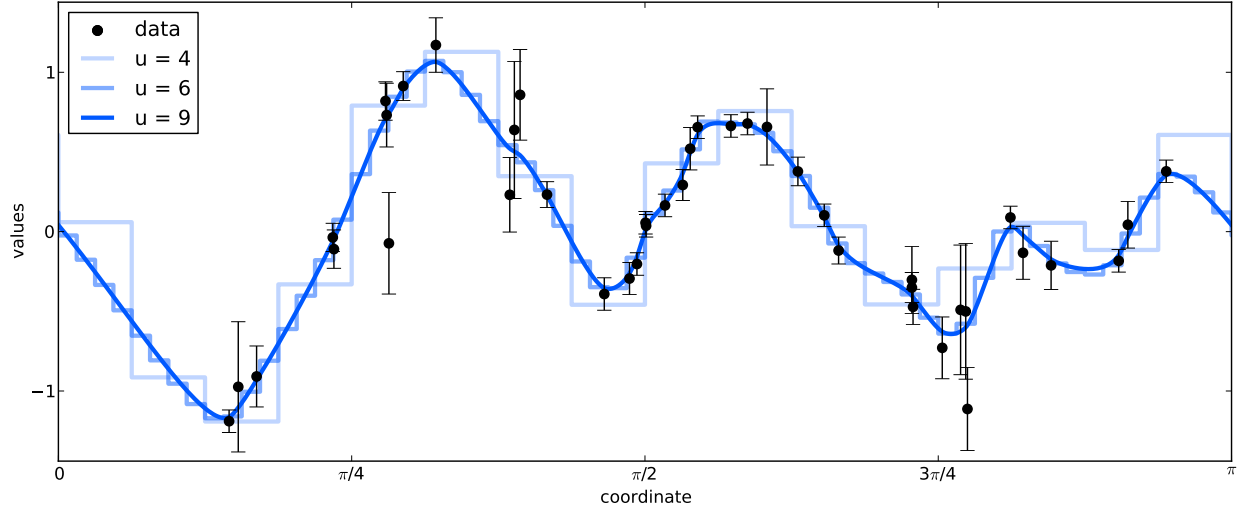


FIGURE 1. 1D signal reconstruction given 42 noisy data points. The (periodic) position space $\mathcal{U} = [0, \pi[$ is discretized by 2^u pixels; i.e., 16, 64, and 512 pixels, respectively.

NIFTY automatizes those normalizations without concerning the user, who simply specifies a space that internally handles all those details. We can choose from a set of preimplemented spaces, which comprises point sets, n -dimensional regular grids, spherical spaces, all their harmonic counterparts, and product spaces constructed as combinations of those. The NIFTY toolkit further offers its users a number of generic and exemplary operators, and several useful numerical tools.

On this basis, NIFTY allows us the abstract formulation and programming of signal inference algorithms. The space \mathcal{U} , on which we define the signal field, is then a mere input for the algorithm. Such an algorithm can run on any space (i.e., regardless of its topology or dimensionality) and with any discretization (i.e., independent of the chosen resolution).

A generic problem is the inference of a (one-dimensional) signal from a set of noisy data points. The signal $s = s(x)$ is a function of position x and not constrained by an analytic model. Fig. 1 illustrates an exemplary set of data points d and the corresponding (symmetric) error bars $\pm\sigma$. The maximum entropy principle thus suggests a zero-mean Gaussian distribution $\mathcal{G}(n, N = \text{diag}[\sigma])$ for the noise n . Given that the data was obtained by a linear measurement process, $d = Rs + n$, described by a response operator R , we find the likelihood $P(d|s)$ to be a Gaussian $\mathcal{G}(d - Rs, N)$.

Next, we need to specify our prior assumptions on the signal; i.e., the individual degrees of freedom s_q in the chosen discretization of the position space.¹ We would *a priori* expect the signal to be continuous and somewhat smooth; i.e., we anticipate a spatial correlation between the degrees of freedom s_q . Again, the maximum entropy principle prompts us to assume a multivariate Gaussian prior $P(s) = \mathcal{G}(s, S)$ in this case. For simplicity, we assume the signal covariance S to be known. However, in real-life applications, where this is seldomly possible, we would resort to more elaborate methods such as critical filtering [7, 8] that allow for the simultaneous reconstruction of s and S from the same data.

Computing the posterior distribution according to Bayes' theorem, we obtain another Gaussian distribution. This reduces the signal reconstruction to the task of maximizing the posterior or, equivalently, minimizing its negative logarithm, which is a quadratic form in the signal s ,

$$-\log P(s|d) = \frac{1}{2}(d - Rs) \bullet N^{-1}(d - Rs) + \frac{1}{2}s \bullet S^{-1}s + \text{const.} \quad (4)$$

$$= \frac{1}{2}s \bullet (R^\dagger N^{-1}R + S^{-1})s - s \bullet (R^\dagger N^{-1}d) + \text{const.}, \quad (5)$$

where \dagger denotes adjunction. The resulting *a posteriori* signal estimate $\langle s \rangle = (R^\dagger N^{-1}R + S^{-1})^{-1}(R^\dagger N^{-1}d)$ is well-known as Wiener filter reconstruction. Notice that this solution, as well as the above equations, involve several

¹ Neglecting any prior assumptions would result in the so-called matched filter that, in this scenario, would yield a signal reconstruction that is a constant line with spikes at the locations of the data points. Although such a reconstruction would satisfy the likelihood, it contradicts our (*a priori* and *a posteriori*) intuition of a good fit.

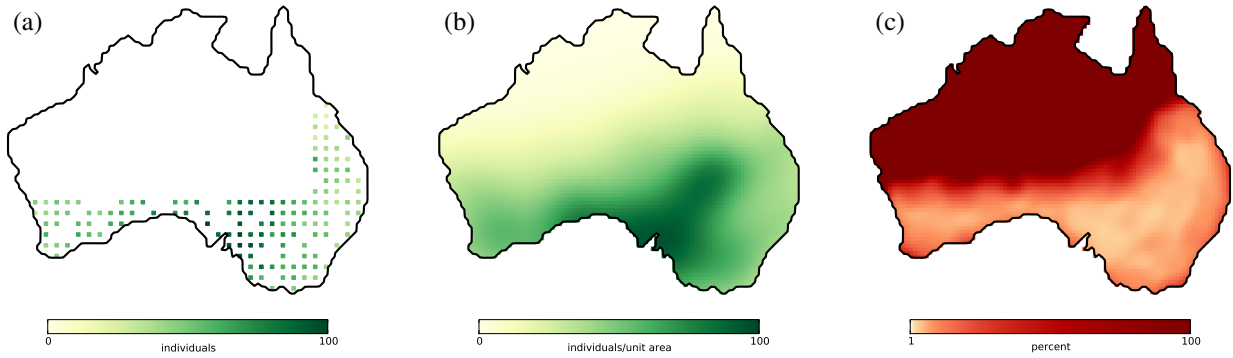


FIGURE 2. 2D signal reconstruction with D³PO “light”: (a) raw data from the census of individuals in a few habitats of unit area, (b) inferred distribution of the species per unit area, (c) relative uncertainty of the inference estimate (cut above 100%).

operations that invoke differently weighted sums according to Eq. (1) or (3), respectively. With NIFTY doing the proper bookkeeping, we can reformulate this filter as an algorithm and apply it regardless of the chosen discretization.² We achieve a resolution independent reconstruction that converges to the continuum limit with increasing resolution. Fig. 1 shows a few signal reconstructions obtained with the same algorithm for different total numbers of pixels.

This functionality has been demonstrated with further numerical examples, which successfully applied Wiener filtering on one- and two-dimensional regular grids, as well as spherical HEALPIX grids [9]. However, Wiener filtering is just one example of the versatile applications one can realize using the NIFTY library. A couple of applications are listed in the next section.

Although NIFTY was originally designed to easily implement algorithms derived in the framework of IFT, it offers a versatile toolkit for any kind of signal inference. Anyone, who wants to infer signal fields from measurement data, can resort to the publically available NIFTY package. NIFTY provides its users an extensive online documentation, tutorials for those new to the subject, and also several demonstration codes. All this (including download and references) can be found on the project homepage at <https://www.mpa-garching.mpg.de/ift/nifty/>.

APPLICATIONS

NIFTY has already been used in a number of applications, some of which are reviewed here to illustrate its versatility. Subsequent to this listing, we discuss the application of the D³PO “light” algorithm to the analysis of census data in detail.

Wiener filtering is a basic application of Bayesian probability theory, in general, and for IFT in particular, since it represents the free theory around which perturbation series can be expanded. The NIFTY package includes a few Wiener filter code examples serving demonstrative and educational purposes [9].

Log-normal statistics are often assumed in order to model physical fields that are strictly positive; e.g., the observable photon flux distribution. NIFTY has been applied to log-normal models that need to solve non-linear, self-consistent filter equations [7, 10, 11, 12].

Spectral smoothness priors for the (Fourier or angular) power spectrum have been studied in the IFT framework [7, 8], and have recently been incorporated into a submodule of the NIFTY package.

D³PO is an inference algorithm for Denoising, Deconvolving, and Decomposing Photon Observations in high energy astronomy and related areas of application [10]. This complex algorithm tackles the problem of inferring the diffuse and point-like photon flux simultaneously from one raw image of photon counts. The raw image is assumed to be corrupted by imprints of the instrument response functions it has been convolved with, and to suffer from Poissonian shot noise. The inference algorithm is derived in the IFT frame and uses a hierarchical Bayesian parameter model for the hyper- and nuisance parameters as well as the signal fields. The denoising (i.e., the suppression of noise) and

² The effect of a bookkeeping error corrupts the reconstruction. An incorrect weighting commonly causes a pixelization dependend suppression or overemphasis of the prior with respect to the likelihood, and thus yields reconstructions that depend on the chosen pixelization.

the deconvolution (i.e., the removal of observational artifacts) are essential steps for signal reconstruction in general. The decomposition task, in which the algorithm allocates flux contributions to the diffuse or point-like component, is the most tricky one. A crucial role plays hereby the exploitation of *a priori* and *a posteriori* correlation structures of the signal fields. We assume log-normal prior for the diffuse signal field, a log-uniform and a smoothness prior for its power spectrum, and inverse-gamma priors for the point-like flux. Given a data set of photon counts, the D³PO algorithm successfully denoises, deconvolves, and decomposes the raw image into a diffuse and a point-like photon flux reconstruction. Currently, photon count data from the INTEGRAL/SPI instrument, the Chandra X-ray observatory, and the Fermi γ -ray space telescope are being analyzed.

RESOLVE (Radio Extended SOURces Lognormal deconVolution Estimator) [11] is a novel algorithm for aperture synthesis imaging of diffuse emission in radio astronomy whose current implementation relies on NIFTY.

Further applications addressing the non-Gaussianity of the cosmic microwave background [13], self-calibration [14], tomography of the Galactic electron density, (weak) gravitational lensing, and computer tomography in the area of medical imaging are in progress.

D³PO “light”

Typical geospatial data in biology is the census of individuals of a certain species within habitats. The counted numbers of individuals within sampled habitats serve as a basis for inferring the species distribution. This is of importance in geo-ecological contexts; e.g., for the (active) conservation of endangered species. For practical reasons, census data is limited to a manageable number of rather small habitats that sample the area of interest sparsely. Fig. 2a illustrates a raw data set from a simulated census of individuals of some arbitrary species (say the “opossum entropia maxima”). The question of how to extrapolate from such a census to unobserved areas in order to obtain an estimate of the species distribution arises. Since ecological systems are highly complex, we adopt a (strongly) simplified view in the following.

Our signal field of interest is the species distribution ρ as a function of (geographic) position. It is convenient to discretize the considered area in cells that are aligned and coextensive with the surveyed habitats. This way, the mapping between data and signal space becomes intuitive. The census data d , on the other hand, is finite and noisy. The geospatial incompleteness of the census can be described by an exposure that is 0 or 1 depending on whether the respective cell has been surveyed or not. Intermediate exposure values might be chosen; e.g., if a habitat was only inspected partly due to inaccessibility, or if there is reason to consider a bias due to nesting or flight behavior of the species. This exposure can be condensed into a diagonal matrix E with which we can describe a deterministic relation between the species distribution ρ and the expected number of counts of individuals λ ,

$$P(\lambda|\rho) = \delta(\lambda - E\rho). \quad (6)$$

The collected census data, however, is subject to noise. The noise statistics can be described by a Poisson distribution, since the measurements is based on individual counts of species members; i.e.,

$$P(d|\lambda) = \prod_i \frac{1}{d_i!} \exp(-\lambda_i + d_i \log \lambda_i). \quad (7)$$

Under those simplistic but adequate assumptions, we can formulate the likelihood of the inverse problem. Its negative logarithm reads,

$$-\log P(d|\rho) = -\log \int D\lambda P(d|\lambda) P(\lambda|\rho) = 1 \bullet E\rho - d \bullet \log(E\rho) + \text{const.}, \quad (8)$$

where 1 is a constant data vector, and the logarithm is applied componentwise. Notice that this likelihood is local; i.e., there is no information cross-talk between different habitats. This is why a prior model for the species distribution is indispensable for the extrapolation into unobserved areas.

In order to construct a prior, we have to consider two crucial constraints. Firstly, the species distribution ρ is strictly positive. Secondly, there are geospatial correlations between the degrees of freedom that, at lowest order, can be described by a covariance. In this case, the maximum entropy principle suggest a multivariate log-normal prior $P(\rho)$; i.e.,

$$-\log P(\rho) = \frac{1}{2}(\log \rho) \bullet S^{-1}(\log \rho) + \frac{1}{2} \log(\det[S]) + \text{const.} \quad (9)$$

This prior model is oversimplified lacking a higher order moments and, more importantly, any ecological constraints involving climate, vegetation, food, and predator information, just to mention a few. More realistic priors could be build up using the maximum relative entropy or Bayesian updating [15] applied to available data applicable to the considered species. However, we want to stress the importance of exploiting spatial correlations and the benefit of *a posteriori* uncertainty information, rather than presenting a realistic model, and for this purpose such a simplistic prior suffices. We choose a convenient parametrization of the *a priori* unknown covariance S that assumes statistical homogeneity and isotropy of the logarithmic species distribution, and further assume Jeffreys and smoothness prior for the introduced hyperparameters according to [7].

Finally, we find a posterior description for the species distribution given census data by combining Eq. (8) and (9). Since this posterior is equivalent to the one derived for denoising, deconvolving, and decomposing photon observations in high energy astronomy, we can apply the D³PO algorithm [10], although in a “light” version without the complications of point-like contributions. Fig. 2b shows the reconstructed species distribution. There, it is evident that the (local) information from the census has been propagated to all positions within the area of interest; i.e., an extrapolation inbetween and beyond the inspected habitats. This became possible by exploiting geospatial correlations encoded in the prior covariance.

At places far away from census habitats, we simply retrieve the prior’s zero-mean. However, knowing the noise statistics and reconstructing the prior covariance, the D³PO algorithm can also provide *a posteriori* uncertainty information. As shown in Fig. 2c, the reconstruction error is around a few percent within the region where the census took place, and increases beyond the census’ borders exceeding 100% at twice to three times the typical distance between inspected habitats. This uncertainty information reveals where we can trust the reconstruction and where additional data is required in order to get more reliable inference results.

ACKNOWLEDGMENTS

My thanks go to my co-developers and collaborators, namely Michael R. Bell, Vanessa Böhm, Sebastian Dorn, Torsten A. Enßlin, Mahsa Ghaempanah, Maksim Greiner, Henrik Junklewitz, Niels Oppermann, Carlos Pachajoa, Martin Reinecke, Gašper Šenk and Philipp Wullstein for the support of the NIFTY project as well as the insightful discussions and productive comments. Furthermore, I thank the “MaxEnt and Bayesian Association of Australia, Inc.” for the awarded student support in order to present this work at the “33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering” in 2013.

REFERENCES

1. T. Bayes, *Philosophical Transactions of the Royal Society* **35**, 370–418 (1763), URL <http://journals.royalsociety.org/content/w24177406t384j06/>.
2. R. T. Cox, *American Journal of Physics* **14**, 1 (1946).
3. E. T. Jaynes, *Physical Reviews* **106 and 108**, 620–630 and 171–190 (1957).
4. T. A. Enßlin, M. Frommert, and F. S. Kitaura, *Phys.Rev.D* **80**, 105005+ (2009), 0806.3474.
5. *Information field theory*, vol. 1553 of *American Institute of Physics Conference Series*, 2013, 1301.2556.
6. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, Technology Press and Wiley, New York, 1949, note: Originally issued in Feb 1942 as a classified Nat. Defense Res. Council Rep.
7. N. Oppermann, M. Selig, M. R. Bell, and T. A. Enßlin (2012), 1210.6866.
8. T. A. Enßlin, and M. Frommert, *Phys.Rev.D* **83**, 105014+ (2011), 1002.2928.
9. M. Selig, M. R. Bell, H. Junklewitz, N. Oppermann, G. M. Reinecke, M., C. Pachajoa, and T. A. Enßlin, *A&A* **554**, A26 (2013), 1301.4499.
10. M. Selig, and T. A. Enßlin (2013), 1311.1888.
11. H. Junklewitz, M. R. Bell, M. Selig, and T. A. Enßlin (2013), 1311.5282.
12. M. Greiner, and T. A. Enßlin (2013), 1312.1354.
13. S. Dorn, N. Oppermann, R. Khatri, M. Selig, and T. A. Enßlin, *Phys. Rev. D* **88**, 103516 (2013), 1307.3884.
14. T. A. Enßlin, H. Junklewitz, L. Winderling, and M. Selig (2013), 1312.1349.
15. *Updating Probabilities*, vol. 872 of *American Institute of Physics Conference Series*, 2006, physics/0608185.